# Search and Analytical Tools

## Topics

Lesson 1 – Search Tools
- Search Engine Tools

Lesson 2 – Search Engines
- Definitions
- Components, Results, Issues
- Individual, Meta
- Google
- Yahoo!
- Other Search Engines

Lesson 3 – Directories
- Definitions
- Yahoo! Directory and Yahoo Search
- Examples

Lesson 4 – Analytical Tools
- SiloBreaker
- Analyst's Notebook
- Factiva Snapshot

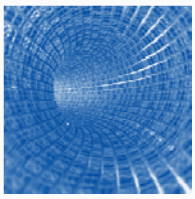
## Objectives

- Introduce and provide information about the various search and analytical tools
- Provide first-hand experience with various search tools

# Lesson 1
## Search Tools

7/19/2012 4

## Introduction to Search Tools



- Internet Search Tools
- How they work
- Sampling of current tools

LESSON 1

7/19/2012 5

## Search Tools

- Web search engines – individual, meta*
- Other search engines
- Directories

**Metasearch engines are search engines that search other search engines and bring you the results as if you had made the effort to search all of those sites independently.*

LESSON 1

7/19/2012 6

## Search Engine Tools

- Search Engines 2 (search engines by topic, region, and function) www.search-engines-2.com
- Best Search Tools Chart from Infopeople.org www.infopeople.org/search/chart.html
- Major Search Engines – Feature Guide http://extremesearcher.com/sechart.pdf
- Search Engines Directory (search engines by topic, general, and specialty) www.searchengineguide.com/searchengines.html
- Choose the Best Search for Your Information Need www.noodletools.com/debbie/literacies/information/5locate/adviceengine.html
- List of Search Engines and Directories http://www.fuld.com/resource-center/intelligence-indexes/

LESSON 1

7/19/2012 7

# Lesson 2

## Search Engines

7/19/2012 8

## Search Engines

A database of Internet content collected in an automated fashion

WIKIPEDIA The Free Encyclopedia

"A web search engine is designed to search for information on the World Wide Web and FTP servers. The search results are generally presented in a list of results often referred to as SERPS, or "search engine results pages". The information may consist of web pages, images, information and other types of files."

(Webopedia)™

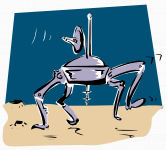
"A program that searches documents for specified keywords and returns a list of the documents where the keywords were found. Although a search engine is really a general class of programs, the term is often used to specifically describe systems like Google, Alta Vista and Excite that enable users to search for documents on the World Wide Web and USENET newsgroups."

LESSON 2

7/19/2012 9

## Search Engine Components

- **Spider/robot/crawler**
  Visits sites on the Web, reads text and links and brings that information back to the search engine index



- **Search Engine Index**
  Where everything the spider finds is indexed
- **Search Engine Software**
  Sifts through the index and brings back matches after the query

LESSON 2

7/19/2012 10

## How Search Engines Work

- Sends out a spider/robot (bot) to the Web
  - Retrieves documents; crawls the Web, capturing text, images, etc.; builds a database
- The indexer
  - Reads these documents; creates an index
- Uses a proprietary algorithm to create its indices
- Some search engines mine data in databases or open directories
- Maintain relative real-time information by continuously running the web crawler

LESSON 2

7/19/2012 11

## Search Engine Results

- Search engines use proprietary algorithms (may include 20 factors) to determine relative relevancy and rank based on items such as:
  - Frequency and location of term
  - Inbound links
  - Popularity - a link to a page is a vote for it
  - Previous search history
- Results of various search engines differ
- Results are affected by the search tools and techniques

LESSON 2

7/19/2012 12

## Search Engine Information Issues

- Not a "real time" web search
  - How current is the information?
- Text dependent
- Barriers to indexing
- Various tools and techniques
- Each search engine has its own algorithm
- Search history affects search results
- Order of the words in your query may affect the ranking of the results
- Use more than one search engine

LESSON 2

7/19/2012 13

## Individual Web Search Engines

- Ask: www.ask.com
- Bing: www.bing.com
- Blekko: www.blekko.com
- Exalead: www.exalead.com/search/
- Google: www.google.com
- HotBot: www.hotbot.com
- Kartoo: www.kartoo.com
- Lycos: www.lycos.com
- Quintura: www.quintura.com
- Yahoo!: www.yahoo.com

LESSON 2

7/19/2012 14

## Metasearch Engines

- 37.com: www.37.com
- Beaucoup: www.beaucoup.com
- Copernic agent: www.copernic.com/en/products/agent
- Dogpile: www.dogpile.com
- Excite: www.excite.com
- Ixquick: www.ixquick.com
- Mamma: www.mamma.com

LESSON 2

7/19/2012 15

## Metasearch Engines

- Metacrawler: www.metacrawler.com
- Search: www.search.com
- Surfwax: www.surfwax.com
- TurboScout: www.turboscout.com
- Yippy: www.yippy.com
- Zuula: www.zuula.com

LESSON 2

7/19/2012 16

## Google - www.google.com

7/19/2012 17

## Google

- One of the most popular
- Estimated web pages indexed – 50 billion (www.worldwidewebsize.com)
- 32 word term limit – applies to search terms and operators but not stop words
- Word stemming
- "Fuzzy" matching
- Cache search capability
- Language search and translation tools
- Limited Boolean capability
- Advanced search options available

7/19/2012 18

## Google Searching
### "ANDing" and "ORing"

"**ANDing**" terms

- It is not necessary to type **and** between the combined words. A space is assumed to be **and**. Advanced Search is an option. Google does not support the use of a plus sign (+) immediately before a word, for searching.

"**ORing**" terms

- **OR** (capitalized) is placed between the combined words. Advanced Search is an option.

7/19/2012 19

## Google Searching
### Excluding terms

Placing a minus sign (-) immediately before a word indicates that you do not want webpages that contain this word to appear in your results. The minus sign should appear immediately before the word and should be preceded with a space.

- Example: anti-virus -software
  - software is excluded
- Example: jaguar -cars -football
  - cars, football are excluded
- Example: -site:bbc.co.uk
  - the bbc.co.uk site is excluded from the search results

7/19/2012 20

## Google Searching
### No stemming

Google does not support the use of a plus sign (+) immediately before a word, for searching. Place double quotes around a single word to stop stemming. The double quotes are also used for exact phrases.

- Example: "ANFO" "explosive" "plan" "c4"
  - will retrieve pages containing: ANFO, explosive, plan, and c4
- Example: ANFO explosive plan c4
  - will retrieve pages containing : ANFO, explosive, explosives, plans, planning, c4, etc.
- Example: "improvised explosive device"
  - will retrieve pages containing: improvised explosive device but not improvised explosive devices, etc.

7/19/2012 21

## Google Searching
### Synonyms/fuzzy searching

Use a tilde (~) before the word that you want Google to generate synonyms for.

- Example: ~nutrition
  - will retrieve: nutrition, eating, health, nutritional, food, protein, calories, etc.

## Hands-on Exercise

## Google Shortcuts

A list of Google advanced search shortcuts is available at http://websearch.about.com/library/cheatsheet/blgooglecheatsheet.htm

**intitle:** Finds pages that have the term(s) in the HTML title element.
Example: intitle:online
Example: intitle:"online strategies"

**allintitle:** Finds pages that have the term(s) in the HTML title element but not necessarily in that order.
Example: allintitle:congress library

**inurl:** Finds pages that have the term somewhere in the URL.
Example: inurl:bbc
Example: inurl:tax (Try this. Are there pages from the IRS website? Why or why not?)

**allinurl:** Finds pages that have the terms somewhere in the URL.
Example: allinurl:google faq

NOTE: **inurl: google inurl:faq** is the same as **allurl:google faq**

**cache:** Displays Google's cached version of a web page.
Example: cache:www.whitehouse.gov

## Google Shortcuts

**link:** Finds pages which contain hypertext links to the exact specified URL.
Example: link:dtic.mil

**site:** Finds pages from the designated web site. It can be used to tell the search engine to only bring back the pages that include your search terms exactly as you typed them, in-order, proximity, etc.
Example: site:bbc.co.uk
Example: "google analytics" site:about.com
Retrieves the pages that contain the phrase, google analytics within the web site, www.about.com
Example: cheese site:.edu
Retrieves the pages that contain the word, cheese, in .edu web sites.

**filetype:** Finds documents of a specific type.
Example: filetype:pdf

**related:** Finds other pages similar to a particular page.
Example: related:howstuffworks.com

**numrange:** Finds a range of numbers.
Example: numrange:1999..2005 (selects the years 1999-2005)
Example: 1999..2005 (selects the years 1999-2005)

7/19/2012 25

## Hands-on Exercise

- Use the various search features of Google; including advanced search.
- Look at the search results from the example above using, site. See what happens when you use site:www.about.com "google analytics."
- A quiz/question: What types of files does Google index?

7/19/2012 26

## Yahoo! - www.yahoo.com

7/19/2012 27

## Yahoo!
### Features and Shortcuts

- Features:
  - Boolean
  - Cache
  - Number of keywords – no limit
- Shortcuts
  - link:
    Example: link:www.fas.org
  - site
  - url
  - intitle



## Hands-on Exercise

## Other Search Engines

- Specialize in a specific media
  - Images
  - Videos
  - Blogs
- or non-Web content
  - News
  - NewsGroups, USENET, Discussion Groups
  - Books

## Special Search Engines/Sites

- Blogs
- News
- Books
- Military
- Government
- Scientific Data

7/19/2012 31

## Video Search Engines

- Bing: www.bing.com
- Google Video: http://video.google.com
- Ixquick: www.ixquick.com
- Reuters Business News and Financial News: www.reuters.com/finance
- YouTube: www.youtube.com

7/19/2012 32

## Blog Search Engines

- Blawg Search: blawgsearch.justia.com/
- Bloggdigger: www.blogdigger.com
- Blog-Search: www.blog-search.com
- Blogsearchengine: www.blogsearchengine.com
- Google blog Search: blogsearch.google.com or www.google.com/blogsearch
- Surfwax: www.surfwax.com
- Technorati: www.technorati.com

7/19/2012 33

## News Search Engines

- Bloglines: www.bloglines.com
- EBSCO (databases; aggregator)
- EIN News: www.einnews.com
- Google News: news.google.com
- Google News Timeline: newstimeline.googlelabs.com
- Ice Rocket Blog Search: www.icerocket.com
- Justia News & Featured Cases: news.justia.com
- My Way: www.myway.com
- Newspaperindex: www.newspaperindex.com
- Nexis: www.nexis.com
- Reuters Business & Financial News: www.reuters.com/finance
- Silobreaker: www.silobreaker.com
- Surfwax: www.surfwax.com
- World News: wn.com
- Yahoo! News: news.yahoo.com

7/19/2012 34

## Message Boards and Forums

- Big Boards: www.big-boards.com
- Boardreader: www.boardreader.com
- BoardTracker: www.boardtracker.com
- Delphi Forums: www.delphiforums.com
- Flickr Groups: www.flickr.com/groups
- Forums Search Board Reader: http://boardreader.com
- Google Groups: http://groups.google.com
- Omgili: http://omgili.com
- Yahoo! Groups: http://groups.yahoo.com

7/19/2012 35

## Book Search Engines

- AddAll: www.addall.com
- Alibris: www.alibris.com
- Amazon: www.amazon.com
- BookFinder.com: www.bookfinder.com
- The British Library: http://blpc.bl.uk
- Google Books: http://books.google.com
- Library of Congress catalog: http://catalog.loc.gov
- The Online Books Page: http://digital.library.upenn.edu/books
- Powell's Books: www.powells.com
- Project Gutenberg: www.gutenberg.org
- WorldCat: www.worldcat.org

7/19/2012 36

## Government/Military

- USA.gov: www.usa.gov
- Federal Digital System: www.gpo.gov/fdsys/
- The Library of Congress, Thomas: http://thomas.loc.gov/home/thomas.php
- Congressional Reference Service for the People: http://opencrs.com
- Various military/DOD websites:
  - U.S. Department of Defense: www.defense.gov
  - U.S. Army Combined Arms center: http://usacac.army.mil/CAC2/index.asp

7/19/2012 37

## International

- Yahoo! International: http://everything.yahoo.com/world/en-US
- Escape Artist: www.escapeartist.com
- Yandex: www.yandex.com

7/19/2012 38

# Lesson 3

## Directories

7/19/2012 39

## Directories

- Organized lists or links
- Reviewed and categorized, and maintained by humans.

**PROS**
- Selective
- Categorized
- Easy to browse
- Searchable

**CONS**
- Limited size of database
- Limited search functionality
- Sponsor bias
- Limited content

LESSON 3

7/19/2012 40

## Yahoo! Directory and Yahoo! Search

- Yahoo! Search returns webpages that contain the words that you entered.
- Yahoo! Directory is human-created; it is a maintained collection of websites organized into categories and subcategories.

Use the Yahoo! Directory when you

- want a better understanding of the subject
- are not sure how to describe what you want
- are searching for words with multiple meanings

LESSON 3

7/19/2012 41

## Directories - Examples

- About: www.about.com
- Blawg Search: http://blawgsearch.justia.com/
- CEOExpress: www.ceoexpress.com
- DMOZ: www.dmoz.org
- Excite: www.excite.com
- Fuld & Company Intelligence Indexes: www.fuld.com/resource-center/intelligence-indexes/
- Infomine: http://infomine.ucr.edu/
- Infoplease: www.infoplease.com
- Ipl2: www.ipl.org
- Kartoo: www.kartoo.com has a directory and a search engine
- The Library Spot: www.libraryspot.com
- World Factbook: www.cia.gov/library/publications/the-world-factbook/
- WWW Virtual Library: http://vlib.org
- Yahoo! Directory: http:/dir.yahoo.com

LESSON 3

7/19/2012 42

# Lesson 4

## Analytical Tools

7/19/2012 43

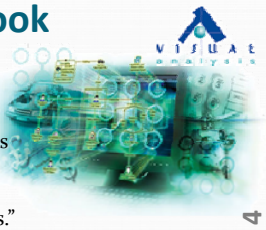
## SiloBreaker - www.silobreaker.com

- "Single system for managing work-flow from:
  - external/internal content aggregation
  - Indexing
  - Mining
  - classification and storage
    - to search, analysis
    - Visualization
    - quality assurance
    - Collaboration
    - decision support and report generation."
- "You can define monitoring targets, trigger content aggregation, perform analyses, and display results in customized dashboards, e-mail alerts or by auto-generated reports."

LESSON 4

7/19/2012 44

## Analyst's Notebook




- Visual Analysis "analysis and visualization product that allows analysts to quickly turn large volumes of data into actionable intelligence with powerful and rich analysis capabilities."
- Images at www.visualanalysis.com/products/anb.php
- Example charts at www.visualanalysis.com/products/anb.php

LESSON 4

7/19/2012 45

## Factiva Snapshot

- Factiva is a product of the Dow Jones Company that offers a collection of over 35,000 global news and information sources from 200 countries in 26 languages.
- It's analytical tool, Factiva Snapshot, filters the news and information that you are monitoring into modules that can be read and scanned.
- Using Factiva Snapshot is available at: www.youtube.com/watch?v=DHmVp8W5qLo&feature=BFa&list=PL804D90BC8AB1F7A9&lf=results_main
- Resource guides are available at:
  - www.factiva.com/en/cp/collateral/files/factiva_snapshot_QRC.pdf
  - www.factiva.com/en/cp/collateral/files/factiva_snapshot_SharePub_QRC.pdf

LESSON 4

7/19/2012 46

## Summary

- Know and use more than one tool!
- Use the right tool with the appropriate search features.
- Search tools are constantly changing. Read the "help" or "about" sections to keep abreast of any changes.

7/19/2012 47